DOCUMENT RESUME

ED 054 192 TM 000 753

AUTHOR Giguere, Constance L.; Baker, Eva L.
TITLE The Development of Criterion-Referenced Tests of Self-Concept and Attitude Toward School.
PUB DATE 71
NOTE 14p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Attitude Tests, *Criterion Referenced Tests, Educational Accountability, Educational Objectives, Measurement Techniques, Peer Relationship, *School Environment, *Self Concept, Self Evaluation, *Student Attitudes, Student Teacher Relationship, Test Construction
IDENTIFIERS *Instructional Objectives Exchange, IOX

ABSTRACT

The development of criterion referenced self-concept and attitude toward school objectives and measures by the Instructional Objectives Exchange (IOX) is discussed. (AG)

ED054192

THE DEVELOPMENT OF CRITERION-REFERENCED TESTS OF SELF-CONCEPT AND ATTITUDE TOWARD SCHOOL

Constance L. Giguere
and
Eva L. Baker

University of California
Los Angeles

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

TM 000 753

One of the prevalent objections to the use of operationally stated objectives either in instruction or as the basis for output measures of PPBS and other accountability schemes is the difficulty encountered in devising objectives and measures to assess the more important goals of education. Relatively little trouble is experienced in writing objectives and items to measure lower level cognitive tasks (Bloom, 1956), but the production of measures for specifically stated higher-level cognitive and affective objectives perplexes most practicioners. The Instructional Objectives Exchange, (IOX), a non-profit educational institution founded with assistance from UCLA's Center for the Study of Evaluation, contracted[1] during the Spring of 1970 to produce objectives and measures in two areas of the affective domain for use in state-wide Title III needs assessment programs. The two of interest were self-concept and attitude toward school.

DEVELOPMENT PROCEDURE

Procedures employed to develop tests of attitude at IOX differ from traditional strategies in that an objectives-based rather than a construct validation approach is employed. Thus, the measures developed do not purport to assess general traits but to provide information of a more specialized nature. The validity of such measures does not depend on a series of highly related correlations but rather on the user's acceptance of the specified premises upon which the instruments are based. Such a design-orientation enables the formulation of instruments appropriate for the evaluation questions in that the measures can relate to dimensions which can be modified as a consequence

[1] The measures described herein were developed through the combined support of a number of state Title III programs (ESEA, 1965).

of changes in school program. Traditional attitude measures are often designed for stability across conditions rather than for sensitivity to change.

TYPES OF MEASURES PLANNED

At the outset, objectives and measures were planned in the broad areas of attitude toward school and self-concept for three grade ranges: Primary (K-3), Intermediate (4-6), and Secondary (7-12) levels. It was anticipated that certain objectives would be tested longitudinally, that is, appear in each grade range, while other measures might concentrate on experiences relatively unique to the age of the learners at a particular level, e.g., dropouts at the secondary level.

In addition, two major types of instruments were planned for use within each grade range, self-report measures and observational measures. Within the self-report category, measures might either be direct (in which the device would possess high content validity and be relatively susceptible to faking) or inferential (in which the intent of the measure was somewhat obscured, resulting in items less fakeable but more dependent upon theoretically-based interpretation).

DEVELOPMENT: SELF-CONCEPT MEASURES

Considerable study of previous writings in the field of self-concept measurement was undertaken by the IOX staff, particularly the work of Wylie (1961), Coopersmith (1967), and Purkey (1970). A survey of extant measures of self-concept was conducted, in which it was determined that most available measures were of the direct self-report variety. While we had originally

planned to derive a fairly global estimate of a general self-concept dimension, an analysis of the literature suggested that a more particularized treatment would be appropriate. Thus, the following four dimensions were selected: 1) Family, i.e., one's self-esteem derived from family interactions; 2) Peer, i.e., one's self-esteem derived from peer relations; 3) Scholastic, i.e. one's self-esteem derived from school interactions; 4) General, i.e., a comprehensive estimate of self-worth. Direct self-report measures for each of the three grade ranges included items related to each of the four dimensions. Seventeen objectives were identified in the self-concept realm, six atthe primary level, six at the intermediate level, and five at the secondary level.

DEVELOPMENT: ATTITUDE TOWARD SCHOOL MEASURES

IOX staff surveyed extant measures related to attitude toward school but the fund of measures and relevant literature from which we could draw was relatively scant. In considering a topic such as attitude toward school, six dimensions emerged as bearing logical relationship to such a general attitude: 1) Teachers, i.e., the students' feelings about teacher behavior, including such aspects as mode of instruction, authority and control, and the interpersonal relationship of teacher to pupils; 2) School subjects, i.e., the students' differential attitudes toward various commonly-taught school subjects; 3) Learning, i.e., toward the learning experience itself as reflected in intellectual curiosity, willingness to study, voluntarism, interest in problem solving; 4) School social structure and climate, i.e., one's attitude toward his school as a social center, a rule-making and rule enforcing entity, and an extracurricular opportunity system; 5) Peer, i.e., one's feeling

regarding the structure and climate of relationships within the peer group rather than toward particular individuals within the group; 6) General, i.e., one's general orientation toward schooling, independent of a particular school. As in the self-concept measures, the direct self-report measures of attitude toward learning included all identified dimensions. Forty-six objectives were identified in the attitude toward school area, twelve at the primary level, fourteen at the intermediate level, and twenty at the secondary level.

DATA ACQUISITION

Both self-concept and attitude toward school scales were tried out and revised on a developmental basis with small samples. Of primary interest was that the format of the items could be easily understood and administered by relatively untrained personnel. Because our development efforts were to focus on the preparation of measures which could be used to assess the quality of an educational program directed to improve learners' attitudes toward school and/or self-concepts, no attempt was made to defend the validity of a measure for an individual child. Rather, we were concerned that the measure would be appropriate for groups of children. Consequently, when inspecting data from preliminary trials, items, and in some cases, entire measures, which produced no variability were discarded, on the assumption that on matters of self-concept and attitude toward school, some variation would be required to demonstrate the objectives' potential sensitivity to change.

FIELD TESTS

A series of field tests on the instruments were conducted during this academic year. One set of field testing involved classes from fourteen

schools, at the elementary, middle and high school levels and included approximately 1500 students. The purposes of these tests was to provide minimal data on use of measures under ordinary classroom conditions in order to determine requisite support materials to accompany the instruments. Data were also inspected to determine if items showed variability. Correlational data, obtained when sets of scales were administered to the same sample, were inspected and related to the mean performance on given measures. Ten measures, drawn from both Attitude Toward School and Self-Concept collections were correlated; however, the largest number of measures any one sample of subjects responded to was three. Nineteen different combinations of measures were listed, seventeen of which included three measures at a time and two consisted of two administered measures.

A CLOSE LOOK AT THREE SELF-CONCEPT MEASURES

A second, more intensive set of trials focused on three self-concept measures:

1) The Self-Appraisal Inventory-Primary Form (a direct measure with thirty-four items and three scales: peer, scholastic, and general).
2) The Class Play (an inferential measure with twenty items and three scales: peer, scholastic, and general).
3) Television Actors (an inferential measure with twenty items and no scales).

and in addition tested the effect of replicable test administration procedures, via audio-tape, or teacher administration on variability of student performances.

The three tests were administered to ten third grade classes, two hundred

sixty-three students, of mixed socio-economic backgrounds. Test analysis questions considered: 1) Kuder-Richardson 21 coefficients to determine the average correlation obtained from all split-half reliability coefficients, 2) Scale score and total test correlations, 3) intercorrelation of scales, and 4) Item difficulty levels.

The ten classes were randomly assigned to the two treatment conditions. Test dates were arranged so that each class received the three tests in varying order to control for practice effects. For the purposes of this field test, the "Self-Appraisal Inventory" and the "Class Play" did not include the family scales because school administrators wished to omit them. Before school on the days of the teacher-administered tests, the teachers were given copies of the measures. This procedure gave them time to familiarize themselves with the method of administration.

At the appointed time on the date assigned, each class received one group administered measure. The researcher was unobtrusively present at each test administration to ensure that the correct procedures were followed. If the class received the teacher administered treatment, the teacher simply distributed the answer booklets to the class and then group administered the test itself. If the class received the tape administered version, the researcher simply turned on the cassette player and played the recorded administration version which was identical in content to the teacher version. The measures were never referred to as "tests" but always as "games" so that the students would not approach the situation anxiously. Helping to contribute to the game impression was the answer booklet which was color coded for each page and included a different cartoon character for each item.

Test analysis results showed that:

1) Kuder-Richardson 21 coefficients differed depending on the method of administration (see Table 1), but a general trend was established with the "Class Play" having the highest coefficients and the "Self-Appraisal Inventory" and "Television Actors" coming in second and third, respectively.

2) Scale score to total test correlations differed depending on administration, but general trends were established (see Table 2). For the "Self-Appraisal Inventory" the scholastic scale correlated the highest with the total test score. The peer and general scales followed second and third, respectively, regardless of administration method. Contrastingly, the "Class Play"'s peer scale correlated highest with the total test score. The scholastic and general scales followed second and third for the tape administered procedure and third and second for the teacher administered procedure.

3) In order to determine if the scales were measuring the same thing or if, in fact, they were measuring different aspects of whatever the total test was measuring, intercorrelations between scales were computed. The low intercorrelations given in Table 3 indicate that the scales were measuring substantially independent aspects of that which the total test was measuring.

4) Item difficulty (p) levels are indicated by the mean p value for each test in Table 4. Nunnaly (1967) suggested that a p value somewhat removed from .5 (5 (around .7) should be a comfort to test makers because it indicates that the majority of the people were able to understand the item well enough to reject or accept it. As can be noted in Table 4, all mean p values fell

around the .7 mark with the exception of those from "Television Actors" which were .44 and .39.

Administration procedure analysis focused on the hypothesis that the variance of the audio-tape treatment scores would be significantly less ($\alpha = .02$) than that of the teacher administered scores. The hypothesis was tested using the Fmax statistic (see Table 5). Since the computed Fmax value exceeded the tabled value, the hypothesis was accepted.

From the above data certain conclusions and implications should be noted:

1) The Kuder-Richardson 21 data indicate that the three measures are most satisfactory for group decisions. Mehrens and Lehmann (1969) point out that if measures are to be used to make group decisions that a coefficient of about .65 may suffice. All KR-21s exceeded .65 except that of "Television Actors" under the tape administration method which was .55.

2) Scale score and total test score correlations demonstrated that the scholastic scale in the "Self-Appraisal Inventory" contributed most to the total test score. For the "Class Play," the peer scale did the same.

3) Low scale intercorrelations showed that each scale measured different aspects of self-concept.

4) All p values fell around the .7 level except those from "Televison Actors" which were .44 and .39.

5) That the main hypothesis, i.e., that the variance of the audio-tape treatment scores would be significantly less than the variance of the teacher treatment scores for each test, was significant has important implications. Clearly, a replicable method of test administration not subject to teacher mood or preparation will provide more representative data than one that

varies from test occasion to test occasion.

SUMMARY

Hopefully, in the development of objective-based tests, certain procedural guidelines will emerge to aid in interpreting data for the purpose of improving measures. Until such techniques are developed, the use of an inferential strategy, i.e., inspecting data for obvious aberrances will necessarily suffice.

TABLES

Table 1
Kuder-Richardson 21 Coefficients by Method and Test

	Teacher	Tape
Self-Appraisal Inventory	.79	.71
Class Play	.82	.73
Television Actors	.73	.55

Table 2
Correlation of Scale Scores
And Total Test Scores

Teacher Administered

Self-Appraisal Inventory:

Scale	Correlation
peer	.81
scholastic	.83
general	.78

Class Play:

Scale	Correlation
peer	.94
scholastic	.81
general	.82

Tape Administered

Scale	Correlation
peer	.78
scholastic	.81
general	.69

Scale	Correlation
peer	.89
scholastic	.77
general	.73

TABLES (cont'd)

Table 3
Correlation Matrix of Scales

Teacher Administered

Self-Appraisal Inventory:

	scholastic	general
peer	.47	.32
scholastic		.38

Class Play:

	scholastic	general
peer	.64	.47
scholastic		.49

Tape Administered

	scholastic	general
peer	.45	.45
scholastic		.55

	scholastic	general
peer	.48	.49
scholastic		.45

Table 4
Mean Item Difficulty Levels

Teacher Administered	Tape Administered
Self-Appraisal Inventory:	
.66	.69
Class Play:	
.76	.82
Television Actors:	
.44	.39

Table 5
Fmax Statistics

Self-Concept Scores	
s^2 for the teacher administered group	32.971
s^2 for the tape administered group	23.490
$F_{max} = \frac{32.971_{(127)}}{23.490_{(136)}}$	= 1.404
Anxiety	
s^2 for the teacher administered group	15.489 (127)
s^2 for the tape administered group	9.322 (136)
$F_{max} = \frac{15.489_{(127)}}{9.322_{(136)}}$	= 1.661*
Extraversion Scores	
s^2 for the teacher administered group	16.039
s^2 for the tape administered group	9.877
$F_{max} = \frac{16.039_{(127)}}{9.877_{(136)}}$	= 1.623*

* significant at the .05 level

BIBLIOGRAPHY

Bloom, Benjamin S. Taxonomy of Educational Objectives (New York: David McKay Company, Inc., 1956).

Coopersmith, S.A. The Antecedents of Self-Esteem (San Francisco: W.H. Freemman Co., 1967).

Mehrens, William A. and Lehmann, Irvin J. Standardized Tests in Education (New York: Holt, Rinehart, and Winston, 1969).

Nunnaly, Jum C. Psychometric Theory (New York: McGraw-Hill Book Company, 1967).

Purkey, William W. The Self Concept and School Achievment (Englewood Cliffs, New Jersey, 1970).

Wylie, Ruth C. The Self-Concept (Lincoln: University of Nebraska Press, 1961).